What Is Claimed Is:

1. A text-speech mapping method comprising:

obtaining silence segments for incoming speech data;

preprocessing incoming transcript data, wherein the transcript data comprises a written document of the speech data;

finding possible candidate sentence endpoints based on the silence segments;

selecting a best match sentence endpoint based on a forced alignment score; setting a next sentence to begin immediately after the sentence endpoint; and

repeating the finding, selecting and setting processes until all sentences for the incoming speech data are mapped.

2. The method of claim 1, wherein the preprocessing incoming transcript data comprises:

scanning the transcript data;

separating the scanned transcript data into sentences; and

placing each word from the scanned transcript data into a dictionary, if the word is not already in the dictionary.

3. The method of claim 2, wherein each word in the dictionary includes information on the pronunciation and phoneme of the word.

4. The method of claim 1, wherein the finding possible candidate sentence endpoints based on the silence segments comprises:

using a dictionary as a table to map words and tri-phonemes for the transcript data;

generating an acoustic model for the speech data, wherein the acoustic model records acoustic features of each tri-phoneme for words in the speech data; and

determining the similarity of the transcript data features obtained from the dictionary with the acoustic model features using a voice engine to find the possible candidate sentence endpoints.

5. The method of claim 4, wherein the voice engine is a HMM (Hidden Markov Model) voice engine.

6. The method of claim 1, wherein upon completion of mapping each sentence, the method further comprises:

obtaining silence segments for each mapped sentence, the method further including determining word level mapping for each mapped sentence, wherein the word level mapping comprises finding possible candidate word endpoints based on the silence segments;

selecting a best match word endpoint based on a forced alignment score;

setting a next word to begin immediately after the word endpoint; and

repeating the finding, selecting and setting processes until all words for the for the mapped sentence are mapped.

7. The method of claim 1, wherein voice activity detection is used to obtain silence segments for incoming speech data.

8. The method of claim 1, wherein a forced alignment process is used to find possible candidate sentence endpoints based on the silence segments, wherein the forced alignment process further includes selecting the best match sentence endpoint based on the forced alignment score.

9. A text-speech mapping system comprising:

a front end receiver to receive speech data, the front end including an acoustic module to model the speech data, wherein the acoustic module to record features of each tri-phoneme of each word in the speech data; and

a voice engine to receive a transcription of the speech data and to obtain features of each tri-phoneme of each word in the transcription from a dictionary, the voice engine to determine candidate sentence and word endings for aligning the speech data with the transcription of the speech data when performing sentence level mapping and word level mapping, respectively.

10. The system of claim 9, wherein the voice engine comprises a HMM (Hidden Markov Model) voice engine to perform alignment of the speech data with the transcription of the speech data.

11. A text-speech mapping tool comprising:

a front end receiver to receive speech data;

a text preprocessor to receive a transcript of the speech data;

a voice activity detector to determine silence segments representative of candidate sentences for the speech data; and

a forced alignment mechanism to determine the best candidate sentence and to align the best candidate sentences from the speech data with sentences from the transcript of the speech data to provide sentence level mapping.

12. The mapping tool of claim 11, wherein the voice activity detector to determine silence segments representative of candidate words for the speech data; and wherein the forced alignment mechanism to determine the best candidate word and to align the best candidate words from the sentences of the speech data with words from the sentences of the transcript of the speech data to provide word level mapping.

13. The mapping tool of claim 11, wherein the forced alignment mechanism further comprises an HMM (Hidden Markov Model) voice engine, wherein the HMM voice engine is used to determine a forced alignment

score for candidate sentences and candidate words based on the silence segments, wherein the best candidate sentence and the best candidate word is based on the maximum forced alignment score.

14. An apparatus comprising:

an automatic text-speech mapping device, the automatic text-speech mapping device, the automatic text-speech mapping device including a processor and a storage device; and

a machine-readable medium having stored thereon sequences of instructions, which when read by the processor via the storage device, cause the automatic text-speech mapping device to perform sentence level mapping, wherein the instructions to perform sentence level mapping include:

obtaining silence segments for incoming speech data;

separating incoming transcript data into sentences, wherein the transcript data comprises a written document of the speech data;

finding possible candidate sentence endpoints based on the silence segments;

selecting a best match sentence endpoint based on a forced alignment score; setting a next sentence to begin immediately after the sentence endpoint; and

repeating the finding, selecting and setting processes until all sentences for the incoming speech data are mapped.

15. The apparatus of claim 14, wherein the machine-readable medium having stored thereon sequences of instructions, which when read by the processor via the storage device, cause the automatic text-speech mapping device to perform word level mapping, wherein the instructions to perform word level mapping include:

obtaining silence segments for each mapped sentence;

finding possible candidate word endpoints based on the silence segments;

selecting a best match word endpoint based on a forced alignment score;

setting a next word to begin immediately after the word endpoint; and

repeating the finding, selecting and setting processes until all words for the mapped sentence are mapped.

16. An article comprising: a storage medium having a plurality of machine accessible instructions, wherein when the instructions are executed by a processor, the instructions provide for obtaining silence segments for incoming speech data;

preprocessing incoming transcript data, wherein the transcript data comprises a written document of the speech data;

finding possible candidate sentence endpoints based on the silence segments;

selecting a best match sentence endpoint based on a forced alignment score; setting a next sentence to begin immediately after the sentence endpoint; and

repeating the finding, selecting and setting processes until all sentences for the incoming speech data are mapped.

17. The article of claim 16, wherein instructions for preprocessing incoming transcript data comprises instructions for:

scanning the transcript data;

separating the scanned transcript data into sentences; and

placing each word from the scanned transcript data into a dictionary, if the word is not already in the dictionary.

18. The article of claim 17, wherein each word in the dictionary includes information on the pronunciation and phoneme of the word.

19. The article of claim 16, wherein instructions for finding possible candidate sentence endpoints based on the silence segments comprises instructions for:

using a dictionary as a table to map words and tri-phonemes for the transcript data;

generating an acoustic model for the speech data, wherein the acoustic model records acoustic features of each tri-phoneme for words in the speech data; and

determining the similarity of the transcript data features obtained from the dictionary with the acoustic model features using a voice engine to find the possible candidate sentence endpoints.

20. The article of claim 19, wherein the voice engine is a HMM (Hidden Markov Model) voice engine.

21. The article of claim 16, wherein upon completion of mapping each sentence, the article further comprises instructions for:

obtaining silence segments for each mapped sentence, the article further including instructions for determining word level mapping for each mapped sentence, wherein the word level mapping comprises instructions for finding possible candidate word endpoints based on the silence segments;

selecting a best match word endpoint based on a forced alignment score;

setting a next word to begin immediately after the word endpoint; and

repeating the finding, selecting and setting processes until all words for the for the mapped sentence are mapped.

22. The article of claim 16, wherein voice activity detection is used to obtain silence segments for incoming speech data.

23. The article of claim 16, wherein a forced alignment process is used to find possible candidate sentence endpoints based on the silence segments, wherein the forced alignment process further includes instructions for selecting the best match sentence endpoint based on the forced alignment score.